驾驭大数据

TAMING THE BIG DATA TIDAL WAVE

FINDING OPPORTUNITIES IN HUGE DATA STREAMS WITH ADVANCED ANALYTICS

[美] Bill Franks 著

黄海 车皓阳 王悦 等 译　张锦沧 张新宇 张琦 审校

WILEY

人民邮电出版社
POSTS & TELECOM PRESS

[illegible]

WILEY

TAMING THE BIG DATA TIDAL WAVE

FINDING OPPORTUNITIES IN HUGE DATA STREAMS WITH ADVANCED ANALYTICS

Bill Franks

□□□□□□□CIP□□□

□□□□□□□□□□□□□□□□Franks□B.□□□□□□□□□—□□□□□□□□□□□□□
2013.1

ISBN 978-7-115-30480-3

Ⅰ□①□… Ⅱ□①□…②□… Ⅲ□①□□□□ Ⅳ□①TP274

□□□□□□□CIP□□□□2012□□298152□

□□□□□

◆□　　□□□Bill Franks
□　　□ □ □□□ □ □□
□　□ □□□ □□□ □ □
□□□□ □□□
□□□□ □ □
◆□□□□□□□□□□□ □□□□□□□□□□14□
□□ 100061 □□□□ 315@ptpress.com.cn
□□ http://www.ptpress.com.cn
□□□□□□□□□□□□
◆□□□700×1000 1/16
□□□16.75
2013□1□□1□
□□□246□□
2013□1□□□□1□□□
□□□□□□□□ □□□01-2012-7920□
ISBN 978-7-115-30480-3
□□□49.00□
□□□□□□□□010□67132692 □□□□□□□□010□67129223
□□□□□□□010□67171154
□□□□□□□□□□□□□□□0021□

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]BDM[illegible]

—Thomas H. Davenport[illegible]

[illegible]Bill Franks[illegible]

—Stuart Aitken[illegible]dunnhumby[illegible]

[illegible]Bill Franks[illegible]“[illegible]”[illegible]Bill[illegible]

—James Taylor, Decision Management Solutions[illegible]
Dicision Management Systems: A Practical Guide to
Using Business Rules and
Predictive Analytics[illegible]

[illegible]——[illegible]

—Diego Klabjan[illegible]
Master of Science in Analytics Program[illegible]

Bill Franks[illegible]

[illegible]Franks[illegible]

—Richard Maltsbarger[illegible]

[illegible]

[illegible]

[illegible]IT[illegible]

[illegible]Bill Franks[illegible]Teradata[illegible]Teradata[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]5[illegible]”[illegible]Kathy. Koontz[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

IDC[illegible]5[illegible]51.4[illegible]“[illegible]”[illegible]Teradata[illegible]IDA[illegible]Integration[illegible]Discovery[illegible]Action[illegible]

[illegible]

[illegible]5[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]Bill Franks[illegible]

[illegible]

[illegible]Bill Franks[illegible]Teradata[illegible]

[illegible]

Teradata[illegible]

2012[illegible]12[illegible]12[illegible]

[illegible]

[illegible]—[illegible]

[illegible]Bill Franks[illegible]

[illegible]Hadoop[illegible]MapReduce[illegible]“[illegible]BDM[illegible]”[illegible]

[illegible]

[illegible]Hadoop[illegible]MapReduce[illegible]

[illegible]

[illegible]

[illegible]Accenture[illegible]Deloitte[illegible]IBM[illegible]

Bill Franks[illegible]Teradata[illegible]
Teradata[illegible]

[illegible]Teradata[illegible]SAS[illegible]Bill Franks[illegible]Teradata[illegible]Bill Franks[illegible]

[illegible]

■ [illegible]1[illegible]“[illegible]”[illegible]Franks[illegible]

■ [illegible]3[illegible]2[illegible]

■ [illegible]4[illegible]MapReduce[illegible]

■ [illegible]“[illegible]”[illegible]“[illegible]”[illegible]5[illegible]

■ [illegible]6[illegible]R[illegible]

■ [illegible]7[illegible]—[illegible]

■ [illegible]Franks[illegible]

[illegible]Franks[illegible]“[illegible]”

[illegible]—[illegible]“[illegible]”[illegible]

Thomas H. Davenport

[illegible]

[illegible]

[illegible]

[illegible]……[illegible]20[illegible]10[illegible]……

[illegible]Facebook[illegible]……

[illegible]……

[illegible]

[illegible]

[illegible]

“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] [illegible]

[illegible] 10[illegible]

（1） [illegible]

[illegible]

（2） [illegible]

[illegible]

（3） [illegible]

[illegible]9[illegible]

[illegible]

[illegible]3[illegible]

（4） [illegible]

[illegible]MPP[illegible]MapReduce[illegible]

（5） [illegible]

[illegible]

（6） [illegible]

[illegible]

[illegible]

[illegible]3[illegible]

[illegible]

[illegible]7[illegible] [illegible]

[illegible]

[illegible]8[illegible] [illegible]

[illegible]

[illegible]9[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]10[illegible] [illegible]

[illegible]

[illegible]11[illegible] [illegible]

[illegible]

[illegible]1[illegible]
[illegible]

[illegible]——[illegible]

[illegible]

[illegible]

1.1 [illegible]

[illegible]Gartner[illegible]Merv Adrian[illegible]2011[illegible]Teradata Magazine[illegible]“[illegible]”[1][illegible]Mckinsey Global Institute[illegible]2011[illegible]5[illegible]“[illegible]”[2]

[illegible]

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□□□□□□□□□□□□□”□
□□□□□□□□□□□□□

□□

■ □□□□□600□□□□□□□□□□□□□□□□□□□□□

■ Facebook□□□□□□300□□□□□□□□□

■ □□□17□□□□□15□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□[3]

□□□□“□”□□□□□□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□Gartner Group□□□□□□□□□□
□“□”□□□□□□□□□□□□□[4]□□□

□□

1.2 □□□□□“□”□“□□”□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”□□□□□□□□□□□□□1□“□”□□2□“□□”□□3□□□□□□□□□□□4□□“□□□□□□□”□□□□□□□□□□

[illegible]4[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

1.3 [illegible]

[illegible]

[illegible]

[illegible]2[illegible]

[illegible]“[illegible]”[illegible]15[illegible]“[illegible]”[illegible]“[illegible]”[illegible]3[illegible]

[illegible]3[illegible]6[illegible]

[illegible]

[illegible]

1.4

[illegible]“[illegible]”[illegible]7[illegible]8[illegible]9[illegible]

[illegible]

[illegible]

[illegible]4[illegible]5[illegible]6[illegible]

[illegible]

1.5 [illegible]

[illegible]8[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]GPS[illegible]

1.6 [illegible]

[illegible]2012[illegible]

[illegible]

[illegible] 2 [illegible] 3 [illegible]

[illegible]

[illegible]

[illegible]

1.7 [illegible]

[illegible]

[illegible]3[illegible]

[illegible]12[illegible]3[illegible]5[illegible]

[illegible]

[illegible]

[illegible]1-1[illegible]


Raw Web Log Data
96.255.99.55 - - [01/Jun/2010:05:28:07 =0000] ″ GET/origin-log.enquisite.com/d.js?id=a1a3af-ly61645&referrer=http://www.google.com/search?hl=en&q=budget=planner&aq=5&aqi=g10&aql=&oq=budget=&gs_rfai=&location=https://money.strands.com/content/simple-and-free-monthly-budget-planner&ua=Mozilla/4.0(compatible; Msie 7.0; Windows NT 6.0; SLCC1; .NET CLR 2.0050727; .NET CLR 3.0.30618; .NET CLR 3.5.30729; Infopath.2)&pc=pgys63w0xgn102in8ms37wka8quxe74e&sc=cr1kto0wmxqik1wlr9p9weh6yxy8q8sa&r=0.07550191624904945 HTTP/1.1 ″ 200 380 ″ - ″ ″ Mozilla/4.0 (compatible; MSIE 7.0; wINDOWS nt 6.0; SLCC1; .NET CLR 2.0.5.727;.NET CLR 3.0.30618; NET CLR 3.5.30729; InfoPath.2) ″ ″ ac=bd76aad174480000679a044cfda00e005b130000 ″

[illegible]1-1 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1.8 [illegible]

[illegible]

[illegible]70[illegible]80[illegible]20[illegible]30[illegible]95[illegible]100[illegible]

[illegible]

[illegible]

1.9

[illegible]3[illegible]RFID[illegible]

[illegible]RFID[illegible]RFID[illegible]10[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]123456789[illegible]”[illegible]3[illegible]10[illegible]“[illegible]”[illegible]10[illegible]

[illegible]“[illegible]”[illegible]3[illegible]10[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1.10 [illegible]

[illegible]ETL[illegible]ETL[illegible]E[illegible]T[illegible]L[illegible]ETL[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1.11

2

3

3

6

EDW EDW EDW

1.12

1.13

1.14

RFID

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]10[illegible]

[illegible]2[illegible]

[illegible]

[illegible]

[illegible][5][illegible]3[illegible]9[illegible]9[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

2.1 [illegible]

360 360 360 360

RFM

10 15

360

2.1.1

95 5 2.5 2.5 1.7

[illegible]98[illegible]

[illegible]

[illegible]RFM[illegible]

2.1.2 [illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]“[illegible]”[illegible]“[illegible]”[illegible]

2.1.3

360

2.1.4

2-1

2-1

购买	请求帮助
产品浏览	转到另一个链接
添加商品到购物车	发表评论
观看视频	注册在线会议
访问下载资源	执行搜索
读/写评论	更多行为

[illegible]

[illegible]

[illegible]

2.1.5 [illegible]

[illegible] “[illegible]” [illegible]

[illegible] cookie[illegible]ID[illegible]“face-less[illegible]”[illegible]ID[illegible]ID[illegible]

[illegible]

[illegible]

[illegible]

2.2 [illegible]

[illegible]

2.2.1 [illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]PC[illegible]

2.2.2 [illegible]

[illegible]

[illegible]JFK[illegible]LaGuardia[illegible]700[illegible]JFK[illegible]30[illegible]20[illegible]700[illegible]JFK[illegible]200[illegible]2[illegible]

[illegible]10[illegible]220[illegible]6[illegible]200[illegible]4[illegible]20[illegible]20[illegible]

[illegible]

2.2.3

DVD

2.2.4

Facebook Twitter

2.3

[illegible]

[illegible]

2.3.1 [illegible]

[illegible]

[illegible]Smith[illegible]

- ■ [illegible]4[illegible]
- ■ [illegible]5[illegible]25[illegible]
- ■ [illegible]
- ■ [illegible]50000[illegible]
- ■ [illegible]15000[illegible]

[illegible]Smith[illegible]Smith[illegible]CD[illegible]Smith[illegible]Smith[illegible]

- ■ [illegible]5[illegible]
- ■ [illegible]
- ■ [illegible]
- ■ [illegible]15[illegible]30[illegible]2[illegible]

[illegible]Smith[illegible]

[illegible]

□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□
□□□

- ■ □□□□□□□□
- ■ □□□□□□□□
- ■ □□□□□□
- ■ □□□□□□□□□

□□
□□

- ■ □□□□□□□
- ■ □□□□□□□□□□
- ■ □□□□□□□□

□□□□□□□□□□□□□□□□□□

2.3.2 □□□□

□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□□□□

□□
□□□□□□□□□□□□□□□□□□□Smith□□□101□□□□□□□□□□Google□
□□□□□□“□□□□□101□□□□□□□□”□□□□□□□□□□□□101□□□□□□
□□□
□□□□

□□□□□□□□□□□□□□□□□□□Smith□□□□□□□□□□□□□□□□□□□□□□□□
□□□

[illegible]Smith[illegible]101[illegible]
Smith[illegible]

[illegible]

2.3.3 [illegible]

[illegible]

[illegible]

[illegible]4[illegible]4[illegible]

- [illegible]90[illegible]
- [illegible]6[illegible]
- [illegible]200[illegible]300[illegible]
- [illegible]100000[illegible]150000[illegible]
- [illegible]
- [illegible]

[illegible]0.62[illegible]4[illegible]

[illegible]

■ [illegible]1[illegible]0.54[illegible]

■ [illegible]2[illegible]0.67[illegible]

■ [illegible]3[illegible]0.78[illegible]

■ [illegible]4[illegible]3[illegible]0.86[illegible]

[illegible]4[illegible]

[illegible]“[illegible]”[illegible]

2.3.4 [illegible]

[illegible]

[illegible]

[illegible]

2.3.5

2.4

■ [illegible]ID[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

□3□
□□□□□□□□□□

□□20□□□

□□2□□□9□□□

□2□□□3□□□□□□□□□□10□□□□□□□□□□□□□10□□□□□□□□□□□□□□□10□□

□□□

□□□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□
- □□□□□□RFID□□□□□□
- □□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□

■

3.1

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

- ■ [illegible]
- ■ [illegible]
- ■ [illegible]
- ■ [illegible]
- ■ [illegible]

[illegible]

[illegible]

3.2 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] 6 [illegible]

[illegible]

[illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

3.3 [illegible]

[illegible]GPS[illegible]GPS[illegible]Foursquare[illegible]Google Places[illegible]Facebook Places[illegible]GPS[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] 5:30 [illegible] 5:45 [illegible] 6:00 [illegible] 5 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]10[illegible]

[illegible]

3.4 [illegible]RFID[illegible]

[illegible]RFID[illegible]RFID[illegible]UPC[illegible]RFID[illegible]Model 123[illegible]Model 123[illegible]

RFID[illegible]RFID[illegible]RFID[illegible]

[illegible]RFID[illegible]RFID[illegible]

[illegible] RFID [illegible]

[illegible] RFID [illegible] RFID [illegible]

RFID [illegible] PC [illegible]

RFID [illegible] RFID [illegible] RFID [illegible] RFID [illegible]

[illegible]

RFID [illegible] RFID [illegible] 5 [illegible]

[illegible] RFID [illegible] 5 [illegible]

RFID [illegible] POS [illegible] RFID [illegible]

RFID[illegible]90[illegible]90[illegible]RFID[illegible]

[illegible]

[illegible]RFID[illegible]

RFID[illegible]HR[illegible]RFID[illegible]

RFID[illegible]Web[illegible]RFID[illegible]RFID[illegible]2[illegible]Web[illegible]“[illegible]”[illegible]

RFID[illegible]RFID[illegible]ID[illegible]RFID[illegible]ID[illegible]RFID[illegible]RFID[illegible]

[illegible]RFID[illegible]RFID[illegible]RFID[illegible]RFID[illegible]

3.5 [illegible]

[illegible] 15 [illegible]

[illegible] 60 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

3.6 [illegible]

[illegible]RFID[illegible]RFID[illegible]RFID[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]RFID[illegible]RFID[illegible]

[illegible]RFID[illegible]RFID[illegible]

[illegible]

[illegible]

[illegible]5[illegible]100[illegible]10[illegible]21[illegible]

[illegible]21[illegible]

[illegible]RFID[illegible]

[illegible]RFID[illegible]

[illegible]“[illegible]”[illegible]ID[illegible]ID[illegible]ID[illegible]ID[illegible]

[illegible]

3.7 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] 3 [illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]

[illegible]

[illegible]

[illegible]

3.8

microtransaction 10

3.9

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

LinkedIn Facebook

[illegible]

[illegible]

[illegible]

3.10 [illegible]

[illegible]

■ [illegible]RFID[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]RFID[illegible]

■ [illegible]

■ [illegible]RFID[illegible]

■ [illegible]

■ [illegible]

[illegible]

[illegible]

[1] Merv Adrian, "Big Data", Teradata Magazine, 1:11, www.teradatamagazine.com/v11n01/Features/Big-Data/.

[2] Mckinsey Global Institute, Big Data: The Next Frontier for Innovation, Competition and Productivity, May 2011.

[3] Ibid.

[4] CEO Advisory: "Big Data" Equals Big Opportunity, Gartner, March 31, 2011.

[5] [illegible]Rebecca Bucnis[illegible]Taking Your Analytics Up a Notch by Integrating Clickstream Data[illegible]SAS Global Forum 2011[illegible]

[illegible]4[illegible]
[illegible]

[illegible]

[illegible] Massively Parallel Processing [illegible] MPP [illegible] MapReduce[illegible]

[illegible]5[illegible]6[illegible]

4.1 [illegible]

[illegible]20[illegible]

[illegible]20[illegible]70[illegible]20[illegible]80[illegible]20[illegible]80[illegible]

[illegible]

[illegible]“[illegible]”[illegible]TB[illegible]2000[illegible]TB[illegible]100[illegible]1TB[illegible]2012[illegible]1TB[illegible]PB[illegible]1000[illegible]

[illegible]TB[illegible]PB[illegible]“[illegible]”[illegible]

[illegible]PB[illegible]Exabyte[illegible]Zettabyte[illegible]4-1[illegible]

[illegible]4-1 [illegible]

1 024 个单位的……	……等于 1 个单位的	注释[1]
KM（Kilobyte）	MB	一张音乐 CD 光盘拥有 600MB 数据
MB（Megabyte）	GB	1GB 可以存储的数据量，等于书架上叠起来大概 9 米多高的书籍
GB（Gigabyte）	TB	10TB 可以存储美国国会图书馆的全部信息
TB（Terabyte）	PB	1PB 可以存储的文本，如果打印出来可以装满 2000 万个 4 门的书柜
PB（Petabyte）	EB	5EB 的信息量等于全人类曾经说过的全部词语
EB（Exabyte）	ZB	使用现在最快的宽带，下载 1Zettabyte 的信息需要至少 110 亿年
ZB（Zettabyte）	YB（Yottabyte）	互联网的全部信息量加起来大概是 1Yottabyte

1 [illegible]http://whatsabyte.com[illegible]

4.2 [illegible]

[illegible] 7 [illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible] Extract [illegible] Transform [illegible] Load [illegible]”[illegible] ETL [illegible]

20 [illegible]

[illegible] Relational Database Management System [illegible] RDBMS [illegible]

[illegible]MapReduce[illegible]“MapReduce”[illegible]

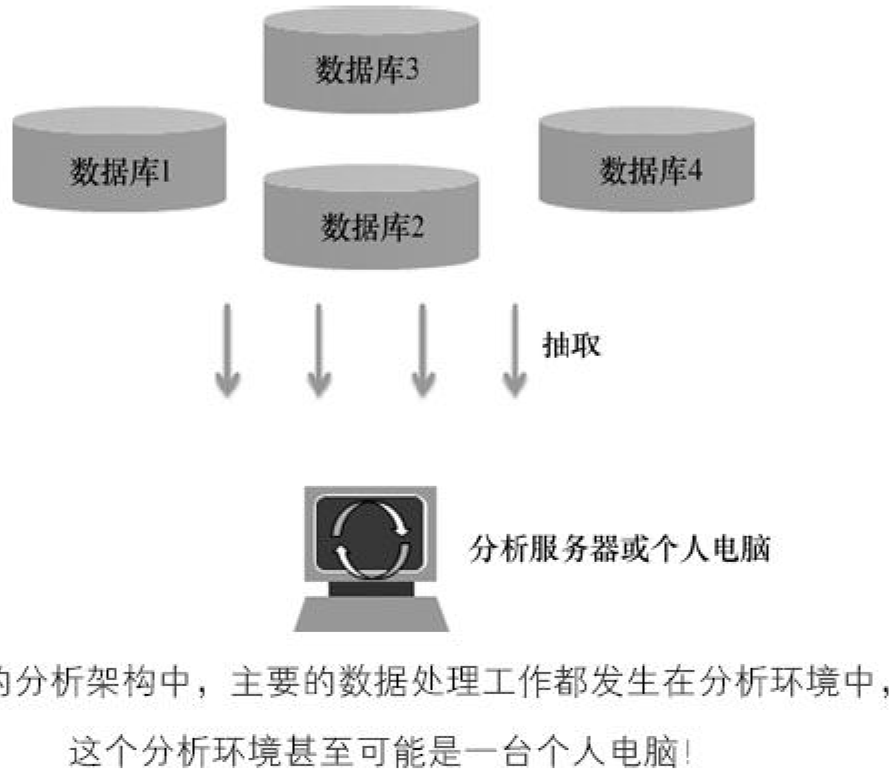
[illegible]

[illegible]

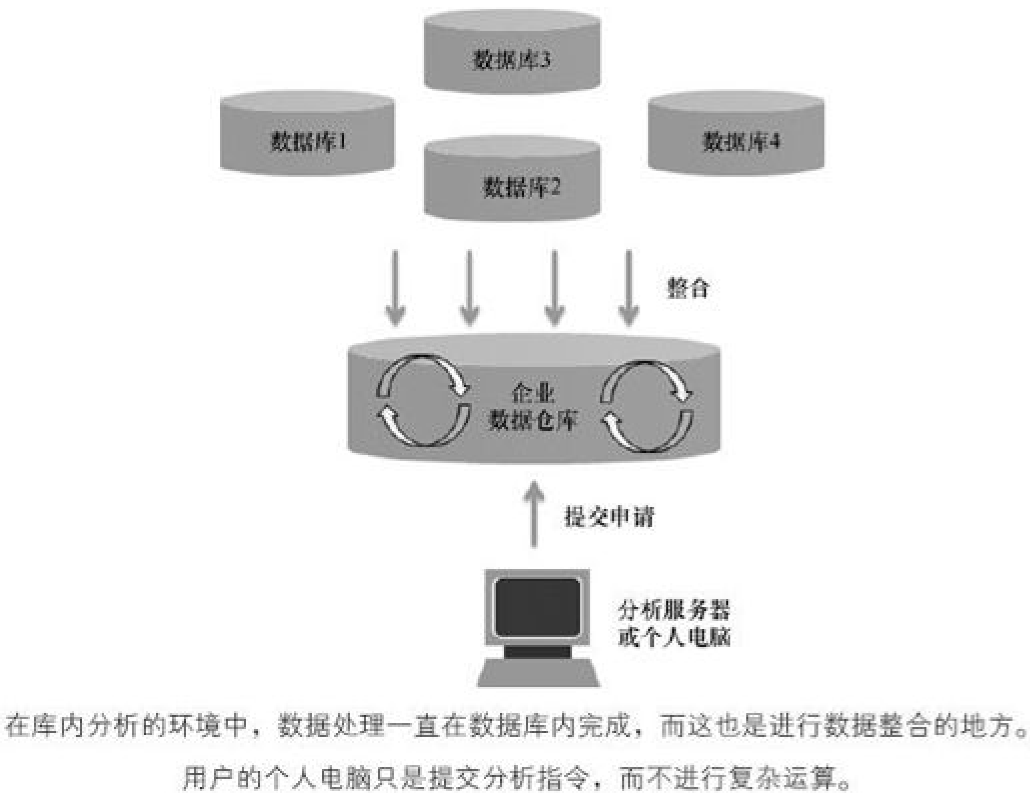
[illegible]“[illegible]Data Mart[illegible]”[illegible]Enterprise Data Warehouse[illegible]EDW[illegible]

[illegible]

[illegible]4-1[illegible]4-2[illegible]


数据库3
数据库1
数据库4
数据库2
抽取
分析服务器或个人电脑
在传统的分析架构中，主要的数据处理工作都发生在分析环境中，
这个分析环境甚至可能是一台个人电脑！

□4-1 □□□□□□□


数据库3
数据库1
数据库4
数据库2
整合
企业
数据仓库
提交申请
分析服务器
或个人电脑
在库内分析的环境中，数据处理一直在数据库内完成，而这也是进行数据整合的地方。
用户的个人电脑只是提交分析指令，而不进行复杂运算。

□4-2 □□□□□□□□□

[illegible]90[illegible]95[illegible]5[illegible]10[illegible]5[illegible]10[illegible]

[illegible]

[illegible]

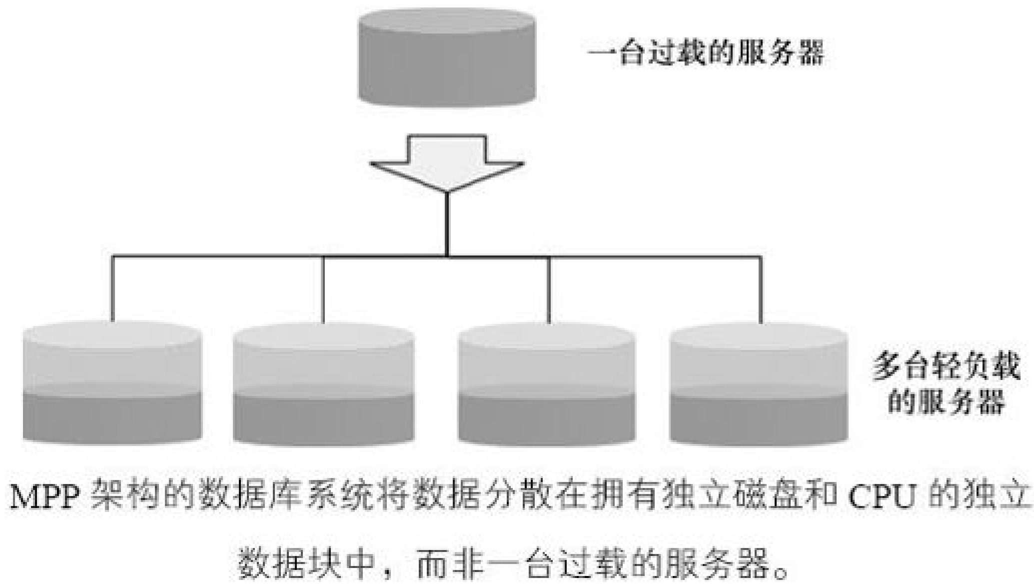
[illegible]20[illegible]90[illegible]Teradata[illegible]

[illegible]

4.3 [illegible]

[illegible]Massively Parallel Processing[illegible]MPP[illegible]MPP[illegible]MPP[illegible]

[illegible]MPP[illegible]CPU[illegible]MPP[illegible]CPU[illegible]MPP[illegible]CPU[illegible]4-3[illegible]MPP[illegible]



[illegible]4-3 [illegible]

[illegible]MPP[illegible]

[illegible]MPP[illegible]MPP[illegible]

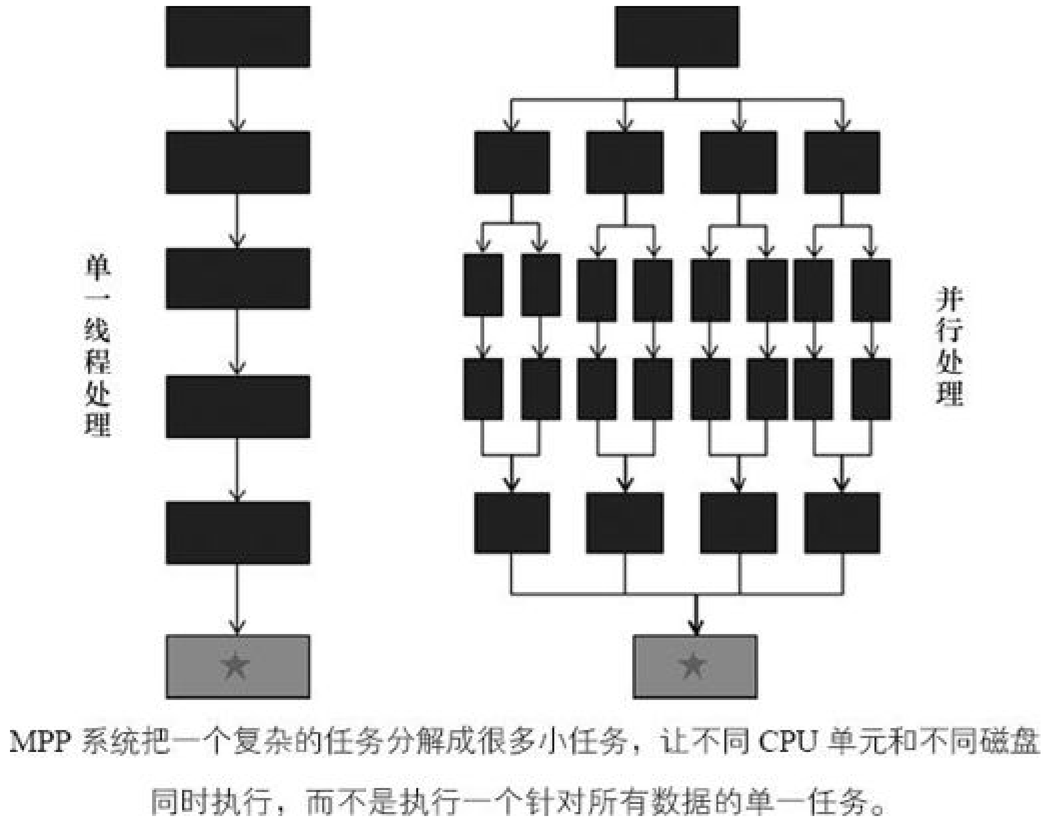
[illegible]1T[illegible]10[illegible]MPP[illegible]1T[illegible]10[illegible]100GB[illegible]10[illegible]100GB[illegible]

[illegible]

[illegible]MPP[illegible]CPU[illegible]

[illegible]20[illegible]10[illegible]100GB[illegible]20[illegible]50GB[illegible]100[illegible]

[illegible]MPP[illegible]4-4[illegible]



[illegible]4-4 [illegible]MPP[illegible]

MPP[illegible]MPP[illegible]CPU[illegible]

4.3.1 [illegible]MPP[illegible]

MPP[illegible]

[illegible]log[illegible]

[illegible]Structured Query Language[illegible]SQL[illegible]SQL[illegible]SQL[illegible]MPP[illegible]SQL[illegible]

[illegible]10[illegible]SQL[illegible]SQL[illegible]SQL[illegible]SQL[illegible]“[illegible]Windowed Aggregates[illegible]”[illegible]SQL[illegible]SQL[illegible]SQL[illegible]

[illegible]SQL[illegible]SQL[illegible]SQL[illegible]

[illegible]SQL[illegible]

[illegible]SQL[illegible]SQL[illegible]

[illegible]MPP[illegible]

[illegible]MPP[illegible]

[illegible]

[illegible]

[illegible]

[illegible]4[illegible]1[illegible]SQL[illegible]2[illegible]UDF[illegible]3[illegible]4[illegible]PMML[illegible]

1[illegible]SQL

SQL[illegible]MPP[illegible]SQL[illegible]SQL[illegible]SQL[illegible]SQL[illegible]SQL[illegible]SQL[illegible]SQL[illegible]

2[illegible]UDF[illegible]

[illegible]User-Defined Function[illegible]UDF[illegible]UDF[illegible]SQL[illegible]UDF[illegible]SQL[illegible]

[illegible]SQL[illegible]

“Select Customer, Sum(sales) …”

[illegible]UDF[illegible]SQL[illegible]

“Select Customer, Attrition_Score…”

[illegible]SQL[illegible]“Attrition_Score”[illegible]SQL[illegible]

UDF[illegible]C++[illegible]Java[illegible]SQL[illegible]SQL[illegible]SQL[illegible]UDF[illegible]UDF[illegible]UDF[illegible]

3[illegible]

[illegible]Embedded Process[illegible]SQL[illegible]

[illegible]

4[illegible]PMML[illegible]

[illegible]Predictive Modeling Markup Language[illegible]PMML[illegible]PMML[illegible]PMML[illegible][1][illegible]PMML[illegible]

[illegible]PMML[illegible]PMML[illegible]

[illegible]

PMML[illegible]PMML[illegible]

[illegible]“SumOfSales”[illegible]PMML[illegible]“SumOfSales”[illegible]

[illegible]PMML[illegible]PMML[illegible]PMML[illegible]PMML[illegible]

PMML[illegible]PMML[illegible]

[illegible]PMML

PMML[illegible]PMML[illegible]2[illegible]

PMML[illegible]PMML[illegible]PMML[illegible]

[illegible]PMML[illegible]PMML[illegible]PMML[illegible]

4.3.2 [illegible]MPP[illegible]

[illegible]MPP[illegible]

[illegible]

[illegible]MPP[illegible]MPP[illegible]

[illegible]MPP[illegible]“[illegible]”[illegible]

[illegible]MPP[illegible]

4.4 [illegible]

[illegible]2009[illegible][2][illegible]3[illegible]

1[illegible]

2[illegible]IT[illegible]

3[illegible]

[illegible]3[illegible]

[illegible]

[illegible]“[illegible]”

[illegible]National Institute of Standards and Technology[illegible]NIST[illegible]5[illegible][3]

1[illegible]

2[illegible]

3[illegible]

4[illegible]

5[illegible]

[illegible]5[illegible]NIST[illegible]NIST[illegible][4]

[illegible]2[illegible]1[illegible]2[illegible]

4.4.1 [illegible]

[illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]“[illegible]”[illegible]“[illegible]”[illegible]

■ [illegible]

[illegible]

[illegible]CPU[illegible]10[illegible]MPP[illegible]MPP[illegible]MPP[illegible]MPP[illegible]

[illegible]

[illegible]

[illegible]

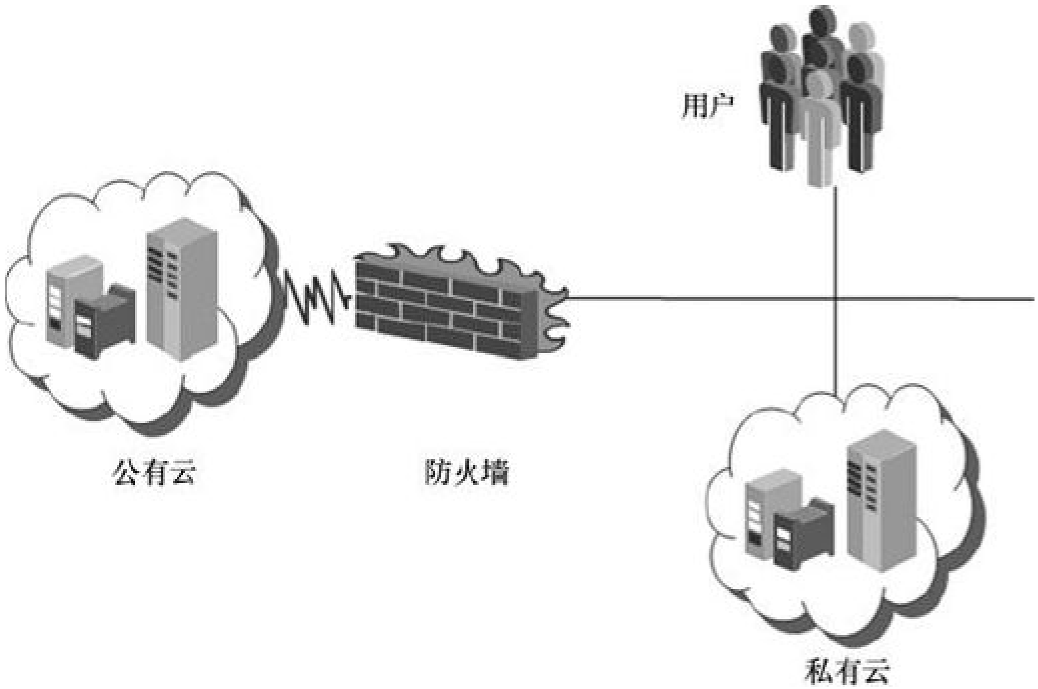
[illegible]

[illegible]

[illegible]

4.4.2 [illegible]

[illegible] 4-5 [illegible]



4-5

[illegible]

4.4.3 [illegible]

[illegible]

4.5 [illegible]

[illegible] SQL [illegible]

[illegible] “[illegible]” [illegible] CPU [illegible]

[illegible]

[illegible]SAS[illegible]

[illegible]6[illegible]

4.6 MapReduce

MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]

MapReduce[illegible]“map”[illegible]“reduce”[illegible]MapReduce[illegible]MapReduce[illegible]MPP[illegible]MPP[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]

[illegible]hashing[illegible]ID[illegible]ID[illegible]

Mapreduce.org[illegible]MapReduce[illegible]Google[illegible]Hadoop[illegible]MapReduce[illegible]Apache[illegible]Hadoop[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]

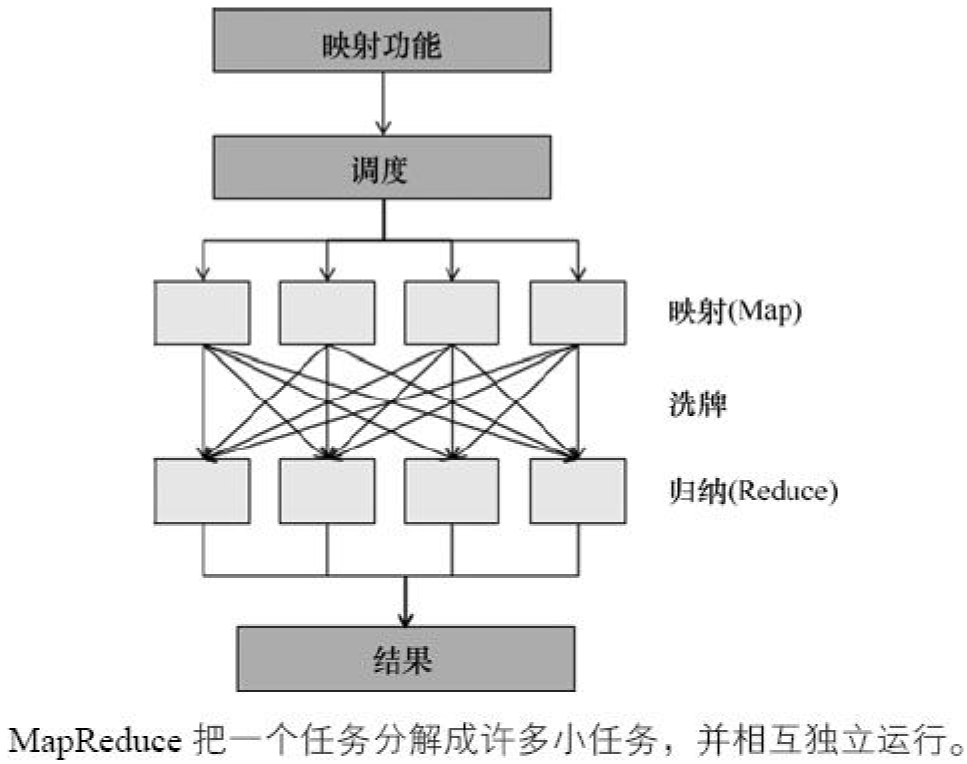
MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]

4.6.1 MapReduce[illegible]

[illegible]20T[illegible]20[illegible]MapReduce[illegible]

[illegible]20[illegible]1T[illegible]MapReduce[illegible]

[illegible]2[illegible]Map[illegible]Reduce[illegible]20[illegible]4-6[illegible]



[illegible]4-6 MapReduce[illegible]

[illegible]“<[illegible]1>[illegible]”“<[illegible]1>[illegible]”“<[illegible]1>[illegible]”[illegible]

[illegible]

MapReduce[illegible]MapReduce[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]26[illegible][5][illegible]A[illegible]B[illegible]C[illegible]

[illegible]“<[illegible]10>[illegible]”“<[illegible]25>[illegible]”“<[illegible]20>[illegible]”[illegible]26[illegible]26[illegible]MapReduce[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]MapReduce[illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]“[illegible]”[illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]RFID[illegible]ID[illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]MapReduce[illegible]

[illegible] MapReduce [illegible] MapReduce [illegible]

4.6.2 MapReduce [illegible]

MapReduce [illegible] MapReduce [illegible]

[illegible] MapReduce [illegible] MapReduce [illegible] MapReduce [illegible] MapReduce [illegible]

[illegible]

[illegible] MapReduce [illegible]

[illegible] MapReduce [illegible]

[illegible] MapReduce [illegible] ETL [illegible] ETL [illegible] ETL [illegible] MapReduce [illegible]

MapReduce MapReduce

MapReduce MapReduce MapReduce MapReduce MapReduce MapReduce

4.6.3 MapReduce

MapReduce MapReduce MapReduce

MapReduce MapReduce MapReduce MapReduce ETL

MapReduce MapReduce MapReduce

[illegible]

4.7 [illegible]

[illegible]MapReduce[illegible]3[illegible]

■ [illegible]

■ [illegible]MapReduce[illegible]Teradata[illegible]Aster[illegible]SQL MapReduce[illegible]MapReduce[illegible]SQL[illegible]

■ [illegible]MapReduce[illegible]MapReduce[illegible]

■ MapReduce[illegible]

■ [illegible]MapReduce[illegible]

[illegible]3[illegible]3[illegible]

4.8 [illegible]

[illegible]

■ [illegible]“[illegible]”[illegible]

■ [illegible]

■ [illegible]MPP[illegible]MapReduce[illegible]

■ [illegible]MPP[illegible]SQL[illegible]UDF[illegible]PMML[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ MapReduce[illegible]

■ MapReduce[illegible]

■ [illegible]MapReduce[illegible]3[illegible]

[illegible]5[illegible]
[illegible]

[illegible]4[illegible]3D[illegible]3D[illegible]3D[illegible]

[illegible]

[illegible]“[illegible]”[illegible]MapReduce[illegible]

[illegible]5.2[illegible]Enterprise Analytic Data Set[illegible]EADS[illegible]EADS[illegible]EADS[illegible]

[illegible]5.3[illegible]

[illegible]

5.1 [illegible]

[illegible] 4 [illegible]

“[illegible]” [illegible]

5.1.1 [illegible]

[illegible]

[illegible]

[illegible]

5.1.2 [illegible]

[illegible] IT [illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]IT[illegible]

■ [illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]IT[illegible]

[illegible]IT[illegible]

■ [illegible]IT[illegible]

■ [illegible]

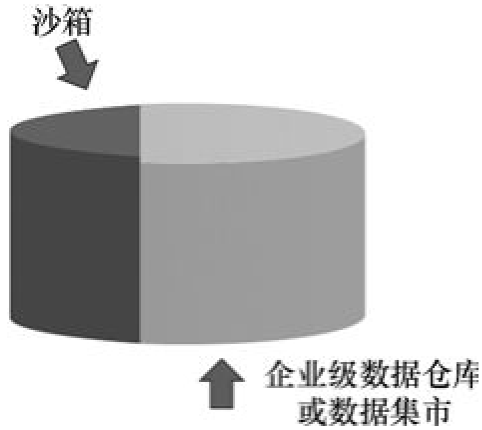
■ [illegible]

■ [illegible]IT[illegible]

■ [illegible]

5.1.3 [illegible]

[illegible]5-1[illegible]

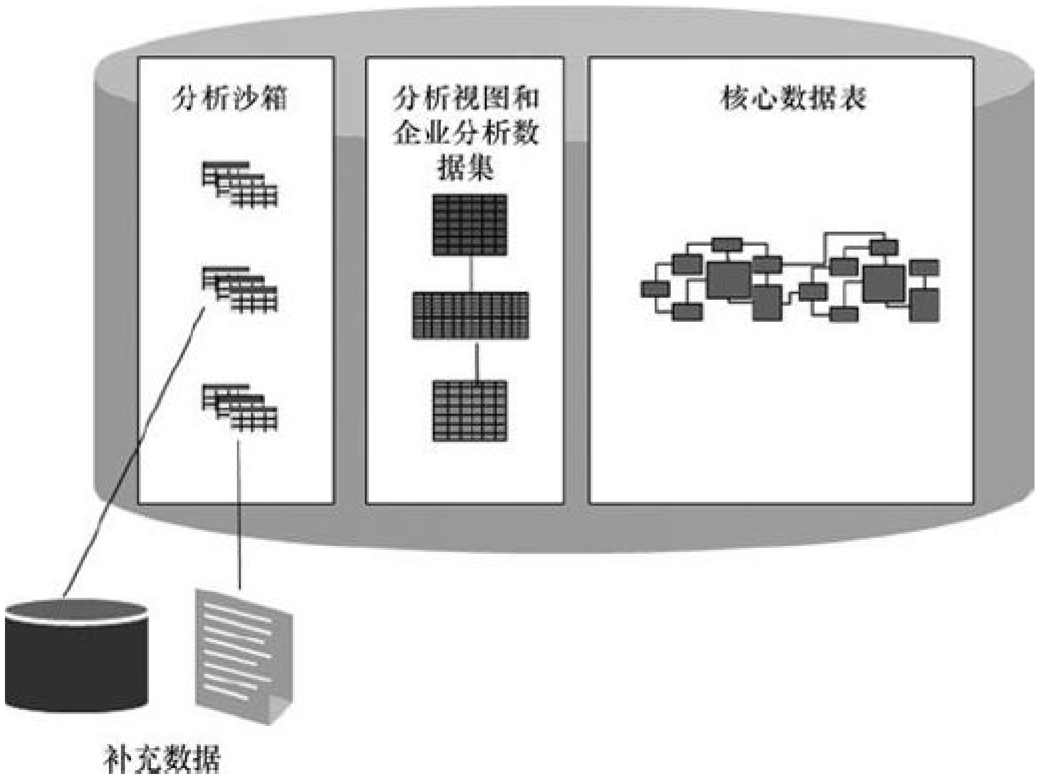


[illegible]5-1 [illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]4[illegible]MapReduce[illegible]

[illegible]

[illegible]5-2[illegible]



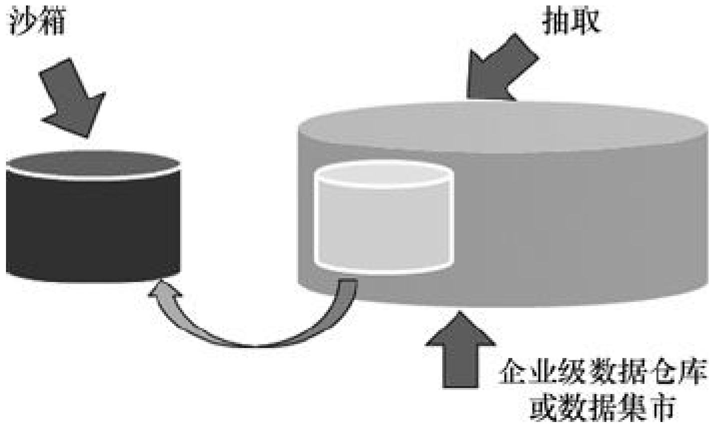
[illegible]5-2 [illegible]

[illegible]MapReduce[illegible]

[illegible]CPU[illegible]

5.1.4 [illegible]

[illegible]5-3[illegible]



[illegible]5-3 [illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

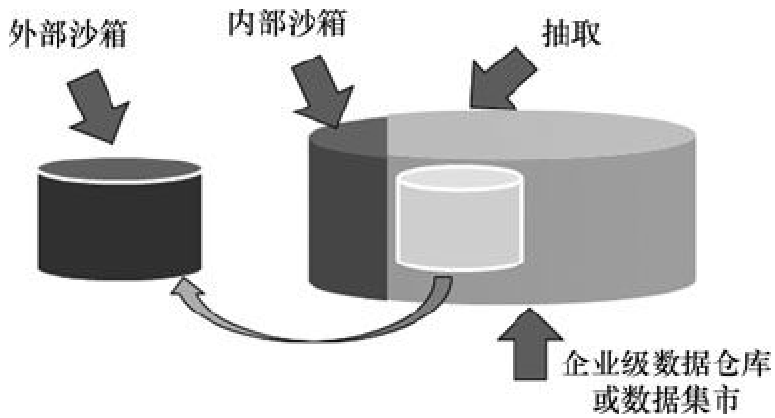
[illegible]

[illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]

5.1.5

5-4



5-4

MapReduce

“

”

[illegible]

[illegible]

[illegible]

[illegible]

5.1.6 [illegible]

[illegible]

[illegible]ETL[illegible]3[illegible]6[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

5.1.7 [illegible]

[illegible]CPU[illegible]10[illegible]100[illegible]

[illegible]5[illegible]

[illegible]

[illegible]5TB[illegible]30TB[illegible]50TB[illegible]5TB[illegible]

[illegible]

[illegible]

[illegible]IT[illegible]

[illegible]“[illegible]”[illegible]1[illegible]2[illegible]IT[illegible]

[illegible]95[illegible]99[illegible]

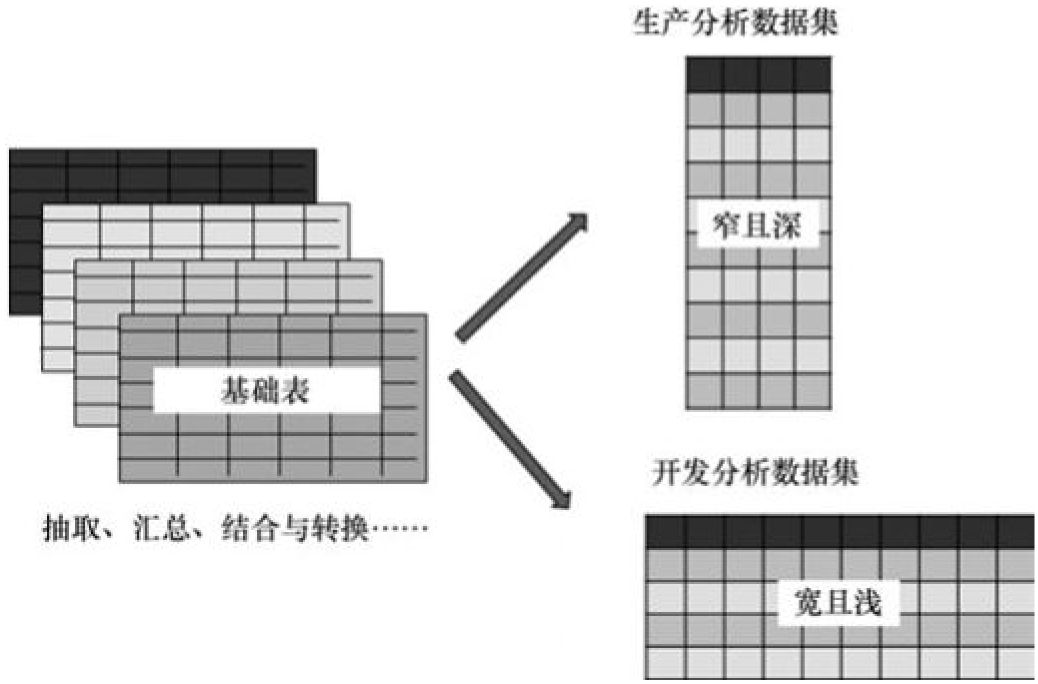
5.2 [illegible]

[illegible]Analytic Data Set[illegible]ADS[illegible]ADS[illegible]

[illegible]“[illegible]”[illegible]

5.2.1 [illegible]

[illegible]5-5[illegible]



[illegible]5-5 [illegible]

[illegible]ADS[illegible]

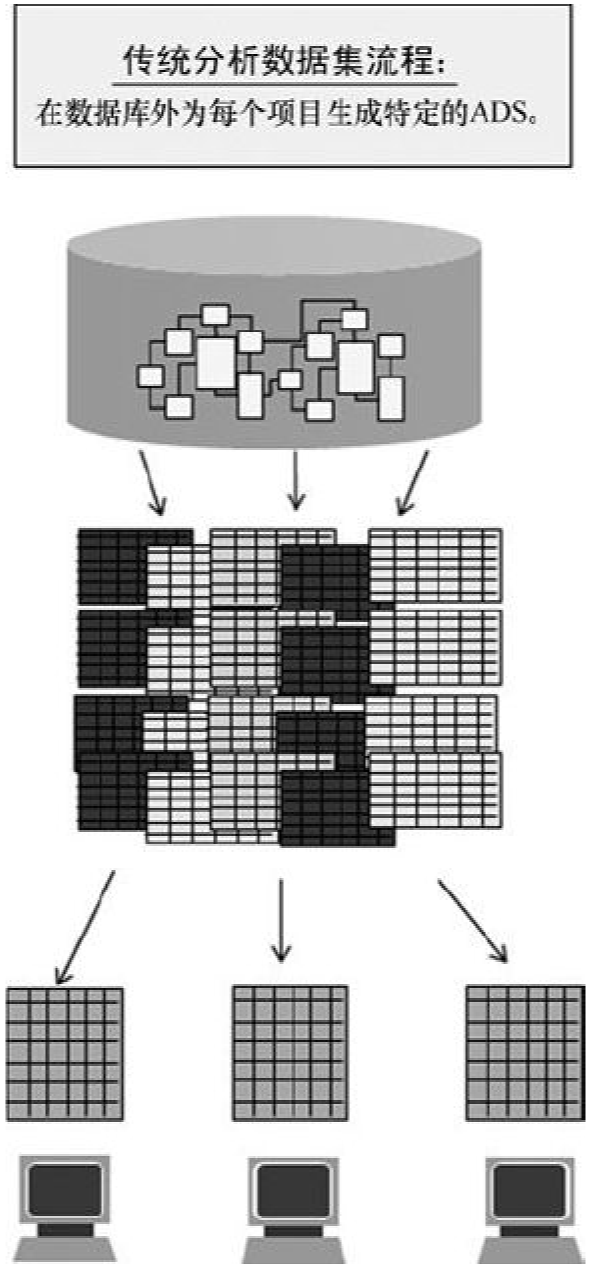
[illegible]

[illegible]500[illegible]10[illegible]12[illegible]3000[illegible]

5.2.2 [illegible]

[illegible]5-6[illegible]

[illegible]ADS[illegible]



[illegible]5-6 [illegible]

[illegible]10[illegible]20[illegible]

[illegible]

[illegible] ADS [illegible]

[illegible]

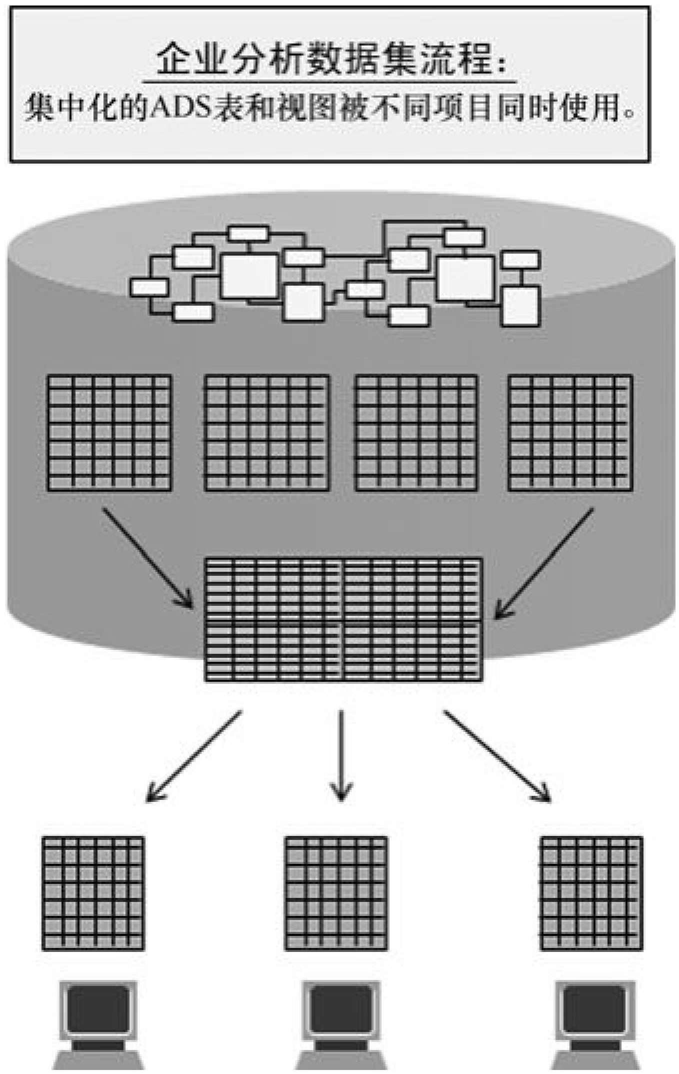
[illegible] ADS [illegible] ADS [illegible] SQL [illegible] UDF [illegible]

5.3 [illegible]

[illegible] EADS [illegible] EADS [illegible]

EADS [illegible] EADS [illegible]

EADS [illegible] EADS [illegible] EADS [illegible] 5-7 [illegible]



[illegible]5-7 [illegible]

EADS[illegible]60[illegible]80[illegible]20[illegible]30[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]

5.3.1 [illegible]

[illegible] EADS [illegible] EADS [illegible] EADS [illegible] EADS [illegible]

[illegible] 6 [illegible] EADS [illegible]

[illegible] IT [illegible] EADS [illegible]

5.3.2 [illegible]

[illegible] EADS [illegible] EADS [illegible]
EADS [illegible]

[illegible]

[illegible] “[illegible]” [illegible]

[illegible] EADS [illegible] 90 [illegible] EADS [illegible]

[illegible]EADS[illegible]

[illegible]EADS[illegible]

[illegible]80[illegible]90[illegible]10[illegible]20[illegible]EADS[illegible]

EADS[illegible]

5.3.3 [illegible]

[illegible]“[illegible]”[illegible]5-8[illegible]EADS[illegible]



[illegible]5-8 EADS[illegible]

[illegible]EADS[illegible]

[illegible]

5.3.4 [illegible]

[illegible]EADS[illegible]

[illegible]

[illegible]EADS[illegible]

5.3.5 [illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

□□□
□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□EADS□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□

□□□
EADS□□□
□□□

□□□□□□□□□□□□□□□□□□□□□EADS□□□□□□□□□□□□□□□□□□
□□□
□□□
□□□

□□□□□□□□

□□□EADS□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□EADS□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□
□□□
□□□□□□□

□□□□□□□EADS□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□
□□□□□□□□□□

□□□
□□□□□□□EADS□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□
□□□□□

5.3.6 [illegible]

[illegible]EADS[illegible]EADS[illegible]EADS[illegible]

[illegible]EADS[illegible]EADS[illegible]CRM[illegible]EADS[illegible]EADS[illegible]CRM[illegible]CRM[illegible]EADS[illegible]CRM[illegible]EADS[illegible]

[illegible]EADS[illegible]EADS[illegible]EADS[illegible]

5.4 [illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

5.4.1 [illegible]

[illegible]CRM[illegible]CRM[illegible]CRM[illegible]ad hoc[illegible]

[illegible]

[illegible]

[illegible]4[illegible]

■ SQL[illegible]
[illegible]SQL[illegible]

■ [illegible]UDF[illegible]
[illegible]

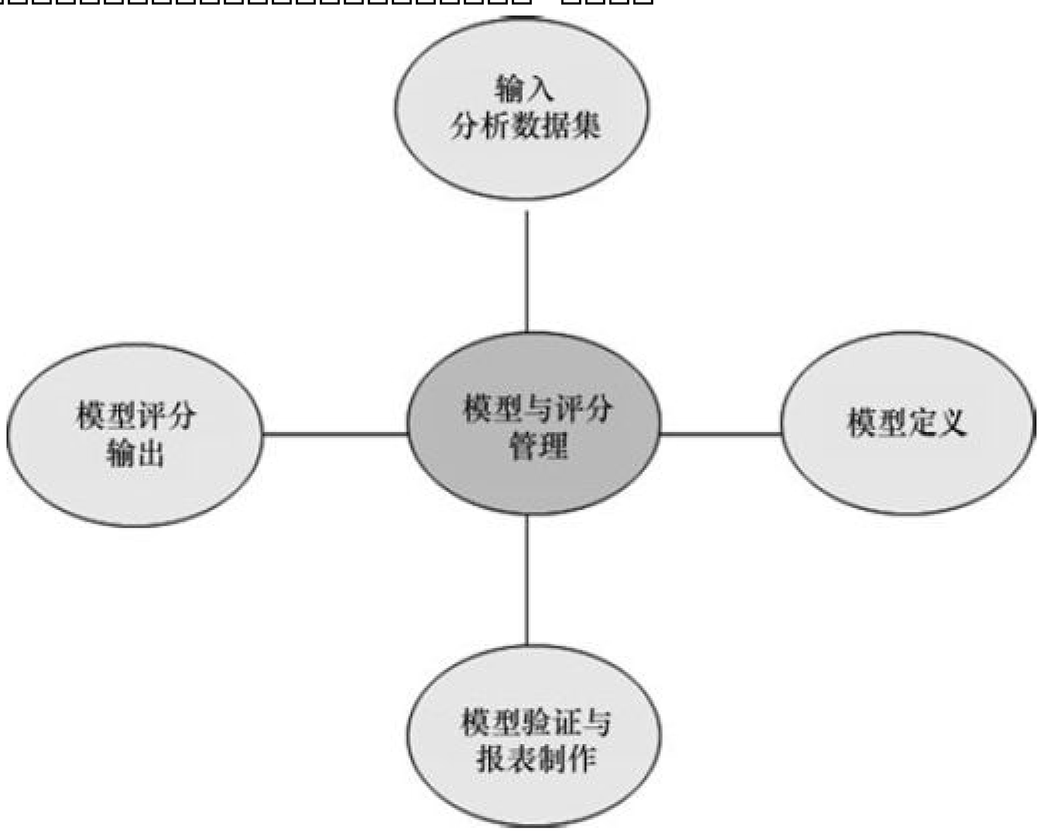
■ [illegible]PMML[illegible]
PMML[illegible]

■ [illegible]

[illegible]4[illegible]4[illegible]
[illegible]

5.4.2 [illegible]

[illegible]4[illegible]5-9[illegible]
[illegible]
[illegible]4[illegible]



[illegible]5-9 [illegible]

1[illegible]

[illegible]EADS[illegible]

■ [illegible]SQL[illegible]UDF[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

2[illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]SQL[illegible]UDF[illegible]

■ [illegible]

3[illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

[illegible]

[illegible]

[illegible]

4[illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

5.5 [illegible]

[illegible]

■ [illegible]

■ [illegible]IT[illegible]

■ [illegible]

■ [illegible] MapReduce[illegible]

■ [illegible]

■ [illegible]ADS[illegible]ADS[illegible]EADS[illegible]

■ EADS[illegible]

■ EADS[illegible]

■ EADS[illegible] EADS[illegible]

■ [illegible]EADS[illegible]

■ [illegible]SQL[illegible]UDF[illegible]PMML[illegible]

■ [illegible]

■ [illegible]4[illegible]

[illegible]6[illegible]
[illegible]

[illegible]

[illegible]4[illegible]5[illegible]

[illegible]

6.1 [illegible]

[illegible]

6.1.1

John Elder Giovanni Seni [6]

James Surowiecki [7]

[illegible]

[illegible]

6.1.2 [illegible]

[illegible]

[illegible] 3000 [illegible] 4000 [illegible] 30 [illegible]

[illegible] 4 [illegible] 5 [illegible]

[illegible]

[illegible] “[illegible]” [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

6.1.3 [illegible]

[illegible]

[illegible] Facebook [illegible] Twitter [illegible]

[illegible] Attensity [illegible] Clarabridge [illegible] SAS [illegible] SPSS [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

6-1

改变重音的位置……	……意思的改变
我没有说比尔的书糟透了。	我没有说，但是我兄弟鲍勃说了！
我**没有**说比尔的书糟透了。	我没有说，你怎么能这样指控我？
我没有**说**比尔的书糟透了。	我没有说，但是我承认我在电子邮件中写了。
我没有说**比尔的**书糟透了。	其他人的书糟透了。
我没有说比尔的**书**糟透了。	我说的不是书，是他的博客糟透了！
我没有说比尔的书**糟透了**。	我仅仅是想表达我不喜欢比尔的书。

6.1.4

[illegible]10[illegible]15[illegible]

[illegible]Google[illegible]Google[illegible]

[illegible]

6.2 [illegible]

[illegible]20[illegible]80[illegible]Job Control Language[illegible]JCL[illegible]

[illegible]PC[illegible]

[illegible]20[illegible]90[illegible]

[illegible] SAS [illegible] SPSS [illegible]

6.2.1 [illegible]

[illegible] 20 [illegible] 90 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] 8 [illegible]

[illegible]

6.2.2 [illegible]

[illegible]

[illegible]SAS[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]1000[illegible]

[illegible]

6.2.3 [illegible]

[illegible]

[illegible]Firefox[illegible]Linux[illegible]Apache Web Server[illegible]

[illegible]Linux[illegible]Apache[illegible]Office[illegible]

[illegible]R[illegible]

[illegible]R[illegible]

[illegible]R[illegible]“R”[illegible]R[illegible]R[illegible]“S”[illegible]S[illegible]R[illegible]S[illegible]R[illegible]R[illegible]Robert Gentleman[illegible]Ross Ihaka[illegible][8]

R[illegible]R[illegible]

[illegible]R[illegible]R[illegible]R[illegible]

R[illegible]R[illegible]C++[illegible]JAVA[illegible]R[illegible]R[illegible]

[illegible]R[illegible]R[illegible]R[illegible]R[illegible]R[illegible]

[illegible]R[illegible]

R[illegible]R[illegible]R[illegible]R[illegible]R[illegible]

[illegible]R[illegible]R[illegible]

R[illegible]R[illegible]R[illegible]R[illegible]

R[illegible]R[illegible]R[illegible]R[illegible]R[illegible]

[illegible]R[illegible]Linux[illegible]Apache[illegible]R[illegible]

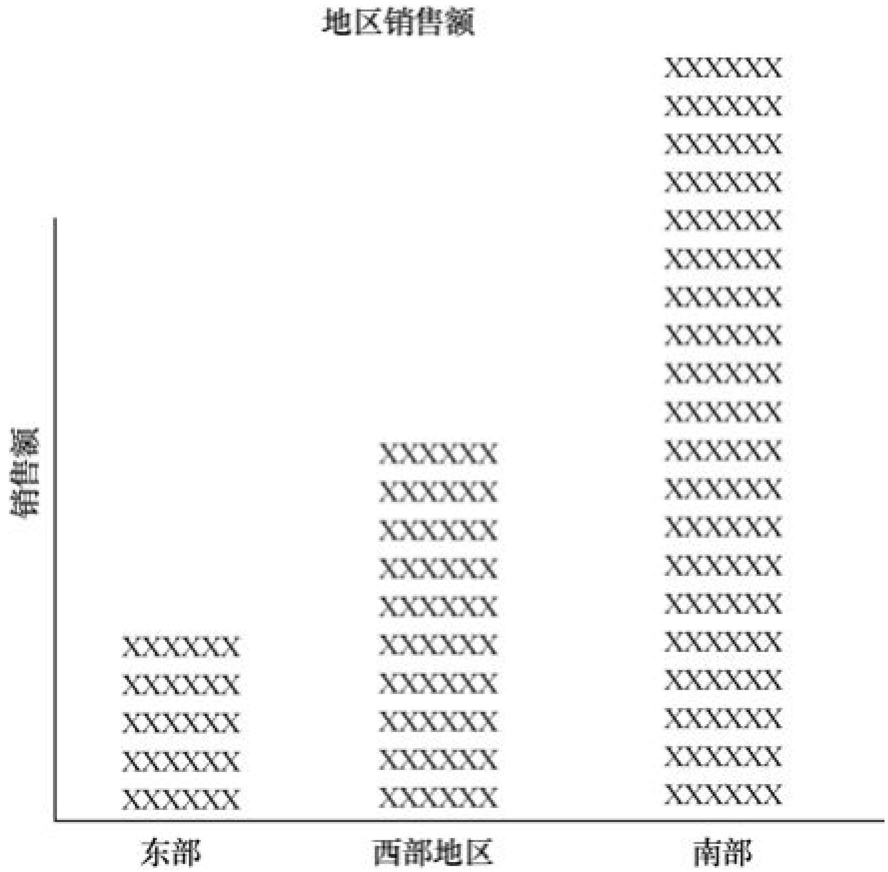
6.2.4 [illegible]

[illegible]Edward Tufte[illegible]A visual Display of Quantitative Information[illegible][9]

Charles Joseph Menard[illegible]1812[illegible][10][illegible]

[illegible]

[illegible]x[illegible]6-1[illegible]“——”[illegible]“|”[illegible]



[illegible]6-1 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]Tableau[illegible]JMP[illegible]Advizor[illegible]Spotfire[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible][11][illegible]3D[illegible]3D[illegible]

[illegible]

[illegible]

[illegible]8[illegible]

3D 2D

Web

6.3 [illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]“[illegible]”[illegible]

■ [illegible]

■ R[illegible]R[illegible]

■ [illegible]

■ [illegible]

[illegible]

[1] [illegible]www.DMG.org[illegible]

[2] [illegible]Clearing the Air on Cloud Computing[illegible]2009[illegible]3[illegible]

[3] [illegible]NIST Working Definition of Cloud Computing[illegible]8-21-09[illegible]15[illegible]http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf[illegible]

[4] [illegible]www.nist.gov/itl/cloud/index.com[illegible]

[5] [illegible]26[illegible]

[6] Giovanni Seni and John Elder, Ensemble methods in Data Mining: Improving Accuracy through Combining Predictions, Morgan and Clay-pool Publishers, 2010.

[7] James Surowiecki, The Wisdom of Crowds, Anchor Books, 2005.

[8] The Comprehensive R Network (CRAN), http://cran.r-project.org/doc/FAQ/R-FAQ.html#Why-is-R-Named-R_03f.

[9] Edward R. Tufte, www.edwardtufte.com/tufte/books_vdqi.

[10] Edward R. Tufte, www.edwardtufte.com/tufte/minard.

[11] ImmerSive Intelligence Colleagues, http://im-tel.org/.

7

7.1

7.1.1

[illegible]BI[illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]

[illegible]prompts[illegible]filters[illegible]

[illegible]IT[illegible]“[illegible]BI[illegible]500[illegible]”

[illegible]“[illegible]”[illegible]IT[illegible]IT[illegible]“[illegible]500[illegible]”[illegible]

[illegible]500[illegible]500[illegible]500[illegible]

[illegible]

[illegible]IT[illegible]

7.1.2

7-1

报表……	分析……
提供数据	提供答案
提供要求的东西	提供需要的东西
标准化	自定义
不涉及人	涉及人
相当不灵活	十分灵活

7.2 G.R.E.A.T

[illegible] G.R.E.A.T [illegible]

7.2.1 [illegible] Guided [illegible]

[illegible]

7.2.2 [illegible] Relevant [illegible]

[illegible]

7.2.3 [illegible] Explainable [illegible]

[illegible]

7.2.4 [illegible] Actionable [illegible]

[illegible]

7.2.5 [illegible] Timely [illegible]

[illegible]

7.3 [illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]SQL[illegible]

[illegible]SQL[illegible]SQL[illegible]—[illegible]SQL[illegible]SQL[illegible]

[illegible]SQL[illegible]

[illegible]

[illegible]SQL[illegible]

7.4 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

7.5 [illegible]

[illegible] B2B [illegible] B2C [illegible] B2B [illegible] B2B [illegible]

[illegible] “[illegible] B2B [illegible]” [illegible]

[illegible]

[illegible]

[illegible]

7.6 [illegible]

[illegible]

[illegible]50[illegible]10[illegible]10[illegible]1024[illegible]10[illegible]99.9[illegible]0.1[illegible]

[illegible]

7.6.1 [illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]7-2[illegible]5[illegible]7-2[illegible]5[illegible]“[illegible]”[illegible]

[illegible]7-2 [illegible]

赛季	汤姆	乔	赢家
1	0.252	0.255	乔
2	0.259	0.266	乔
3	0.237	0.241	乔
4	0.253	0.255	乔
5	0.256	0.257	乔

[illegible]……[illegible]7.2[illegible]5[illegible]

[illegible]

[illegible]

[illegible]

[illegible]7-3[illegible]5[illegible]t[illegible]

[illegible]

[illegible]7-3 [illegible]

年	汤姆：平均击球率	汤姆：击球数	汤姆：得分	乔：平均击球率	乔：击球数	乔：得分	赢家
1	0.252	123	31	0.255	341	87	乔
2	0.259	355	92	0.266	109	29	乔
3	0.237	139	33	0.241	377	91	乔
4	0.253	304	77	0.255	294	75	乔
5	0.256	363	93	0.257	206	53	乔
总计	0.254	1 284	326	0.252	1 327	335	汤姆!!

*[illegible]

[illegible]7-2[illegible]

[illegible]95[illegible]99[illegible]95[illegible]5[illegible]20[illegible]1[illegible]

[illegible]95[illegible]99.9[illegible]

[illegible]1[illegible]A[illegible]B[illegible]

7.6.2 [illegible]

[illegible]100[illegible]

[illegible]

[illegible] 99 [illegible] 10 [illegible] 2 [illegible] 10 [illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

7.7 [illegible]VS[illegible]

[illegible] 10 [illegible]

[illegible] 800 [illegible] 1200 [illegible] 1000 [illegible]

[illegible]

[illegible] A [illegible] B [illegible] A [illegible] 2.5235 [illegible] B [illegible] 2.5237 [illegible] 0.0002 [illegible] “[illegible]”

[illegible]

[illegible] 20 [illegible] 2000 [illegible] 1 [illegible] 1 [illegible] 1 [illegible]

[illegible] N [illegible] 100 [illegible]

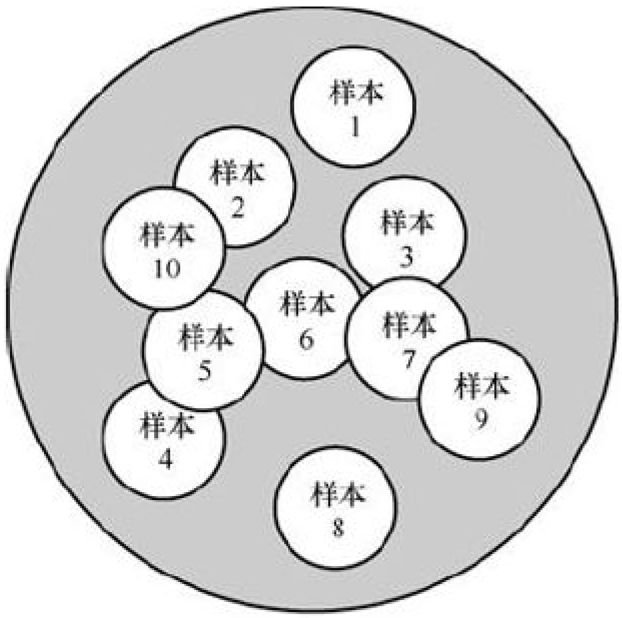
[illegible] 10 [illegible] 10 [illegible]

[illegible]

[illegible] 100 [illegible]

[illegible]

[illegible]



任意给定问题可能都只需要一小部分样本数据。但时，使用样本数据获得解决问题的分析方法后，就需要使用全部数据进行分析了。

[illegible]7-1 [illegible]

7.8 [illegible]

[illegible]

10

10

7.9

■

■

■ G.R.E.A.T

■

■

- ■ [illegible]
- ■ [illegible]
- ■ [illegible]
- ■ [illegible]
- ■ [illegible]

[illegible]8[illegible]
[illegible]

[illegible]7[illegible]3[illegible]5[illegible]

[illegible]

8.1 [illegible]

[illegible]MapReduce[illegible]

[illegible]

[illegible]

8.2

8.3 [illegible]

[illegible]

8.3.1 [illegible]

[illegible]Bart[illegible]Bart[illegible]

[illegible]“[illegible]”[illegible]Bart[illegible]

Bart[illegible]

Bart[illegible]

8.3.2 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] Mark [illegible]
Mark [illegible] Mark [illegible]

[illegible] Mark [illegible] Mark [illegible] Mark [illegible] Mark“[illegible]”[illegible] Mark [illegible]

8.3.3 [illegible]“[illegible]”

[illegible] HR [illegible]“[illegible]”[illegible]

[illegible] HR [illegible]

HR [illegible]“Bill [illegible]”[illegible]“[illegible]

[illegible]”

[illegible]

8.4 [illegible]

[illegible]

8.4.1 [illegible]

[illegible]

[illegible]

8.4.2 [illegible]

[illegible]

[illegible]

[illegible]“@#[illegible]$&*!!!”[illegible]

[illegible]10[illegible]“@#[illegible]$&*!!!”[illegible]

1[illegible]

[illegible]

[illegible]“M”[illegible]“F”[illegible]“U”[illegible]“H”[illegible]10000000[illegible]

[illegible]“H”[illegible]“U”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

2[illegible]

[illegible]

[illegible]

[illegible]1[illegible]

8.4.3 [illegible]

[illegible]

[illegible]Mark[illegible]

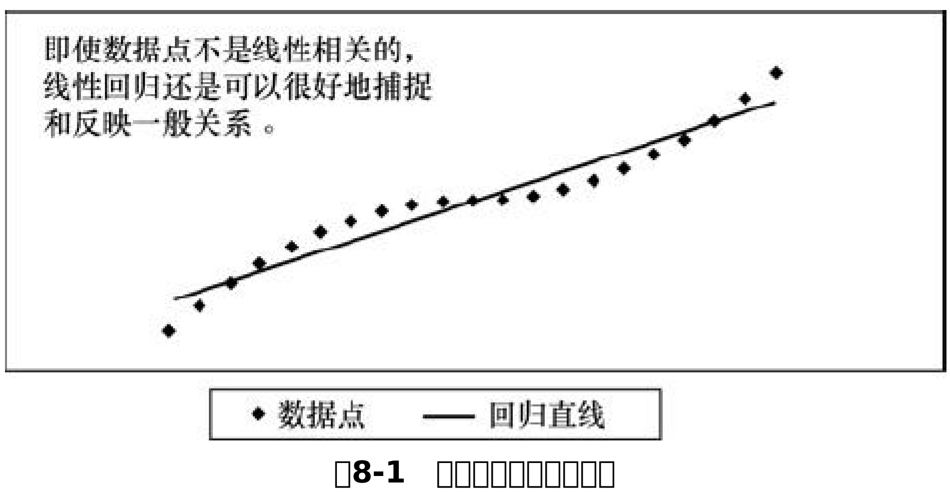
[illegible]

[illegible]IT[illegible]

1[illegible]

[illegible]2[illegible]2[illegible]

[illegible]5.32526[illegible]2[illegible]2[illegible]5.32526[illegible]2[illegible]5[illegible]3[illegible]2[illegible]


即使数据点不是线性相关的，
线性回归还是可以很好地捕捉
和反映一般关系。
数据点
回归直线

[illegible] IT [illegible]

[illegible] 25 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

8.4.4 [illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1[illegible]

[illegible]50[illegible]

[illegible]50[illegible]

[illegible]

[illegible]

[illegible]

[illegible]10[illegible]

2[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

8.4.5 [illegible]

[illegible] A [illegible] B [illegible] C [illegible] D [illegible]

[illegible] Daniel H. Pink [illegible] Riverhead Trade [illegible] 2006 [illegible] A Whole New Mind [illegible] Pink [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

MapReduce

8.5

[illegible]CPA[illegible]CFP[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

8.6 [illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]50[illegible]

■ [illegible]

■ [illegible]

[illegible]9[illegible]
[illegible]

[illegible]

[illegible]

[illegible]

9.1 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

9.2 [illegible]

[illegible] 8 [illegible]

9.3

[illegible]

[illegible]

[illegible]

[illegible]2009[illegible]2011[illegible]HR[illegible]20[illegible]HR[illegible]HR[illegible]

9.4 [illegible]

[illegible]Tom Davenport[illegible]Jeanne Harris[1][illegible]3[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

9.4.1 [illegible]

[illegible]COO[illegible]CMO[illegible]

[illegible]

[illegible]

[illegible]5[illegible]

[illegible]20[illegible]

[illegible]

[illegible]

9.4.2 [illegible]

[illegible] CFO [illegible] COO [illegible] CIO [illegible]

[illegible] 3 [illegible] 3 [illegible]

[illegible] 10 [illegible] 10 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

9.4.3 [illegible]

[illegible]

[illegible] Center of Excellence [illegible] COE [illegible] Center of Expertise [illegible] COE [illegible] COE [illegible] COE [illegible]

[illegible]

[illegible]

9.5 [illegible]

[illegible]

[illegible]

9.5.1 [illegible]

[illegible] “[illegible]” [illegible]

[illegible] Bob [illegible] Sue [illegible] A [illegible] Sue [illegible] Bob [illegible] B [illegible] Bob [illegible] Sue [illegible]

9.5.2

[illegible]

[illegible]

9.6 [illegible]

[illegible][2]

[illegible]6[illegible]SQL[illegible]Join[illegible]SQL[illegible]

[illegible]

[illegible]7[illegible]“[illegible]”[illegible]

[illegible]

9.6.1 [illegible]

[illegible]

Jane[illegible]“[illegible]”Jane[illegible]

John[illegible]CFO[illegible]Jane[illegible]“[illegible]”[illegible]John[illegible]

[illegible]Joe[illegible]Joe[illegible]Joe[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible] 8 [illegible]

9.6.2 [illegible]

[illegible] Barb [illegible] Barb [illegible]

[illegible]

[illegible] 10 [illegible] 10 [illegible]

[illegible]

[illegible]

[illegible]Barb[illegible]Barb[illegible]Barb[illegible]Barb[illegible]Barb[illegible]CRM[illegible]Barb[illegible]

9.7 IT[illegible]

[illegible]IT[illegible]IT[illegible]IT[illegible]

[illegible]9-1[illegible]IT[illegible]IT[illegible]

[illegible]9-1 [illegible]IT[illegible]

分析团队要做的事情	IT 团队要做的事情
大量使用系统资源	严格管理系统资源的使用
创建表格并使用大量的空间	限制创建表格和空间的使用
运行复杂的即席查询	使复杂的即席查询保持在最低水平
打破常规	让用户守规矩
使用新方法做试验	坚持使用正式批准的方法
在规章和约束下工作	执行规章和约束

IT[illegible]

[illegible]IT[illegible]CIO[illegible]CEO[illegible]C[illegible]CEO[illegible]CEO[illegible]

IT[illegible]

[illegible]IT[illegible]

[illegible]

[illegible]IT[illegible]IT[illegible]IT[illegible]

[illegible]4[illegible]5[illegible]IT[illegible]

[illegible]IT[illegible]

[illegible]

[illegible]IT[illegible]IT[illegible]

9.8 [illegible]

[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]IT[illegible]

[illegible]

[1] Tom Davenport[illegible]Jeanne Harris[illegible]Organizing Analysts[illegible]2009[illegible]6[illegible]23[illegible]

[2] [illegible]http://iianalytics.com/category/factulty-blogs/bill-franks/[illegible]

□10□
□□□□□□

□□□□□□□□MerriamWebster□□□□□□□□

1□“□□□□□”□

2□“□□□□□□□□□□”□[□1□]

□□□

□□

□□□

10.1 □□□□□□□

□2008□□2011□□□□□□□□□□□□□□□□“□□□□□”□□

10.2

10.3

10.4

[illegible]

[illegible]

[illegible]

[illegible]100[illegible]

[illegible]

10.5 [illegible]

[illegible]10-1[illegible]

[illegible]10-1 [illegible]

退休计划	风险投资
选择已有的、被广泛认可的公司	选择新的、未被广泛认可的公司
平衡收益和亏损	着眼于收益而不是亏损
混合搭配几种适中的回报	“赚大钱”或者“赔老本”
严格控制波动性	波动性很大
用于退休和日常资产	不用于影响生计的资产

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] 100 [illegible]

[illegible]

10.6 [illegible]

[illegible]

[illegible]

[illegible]9[illegible]

10.6.1 [illegible]1[illegible]

[illegible]4[illegible]MapReduce[illegible]

[illegible]

[illegible]5[illegible]MapReduce[illegible]MapReduce[illegible]

[illegible]

10.6.2 [illegible]2[illegible]

[illegible]

[illegible]

10.6.3 [illegible]3[illegible]

[illegible]

[illegible]

10.6.4 [illegible]4[illegible]

[illegible]IT[illegible]IT[illegible]

[illegible]20[illegible]20[illegible]

10.6.5 [illegible]5[illegible]

[illegible]

[illegible]

10.6.6 [illegible]

[illegible]10[illegible]10[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]8[illegible]

[illegible]

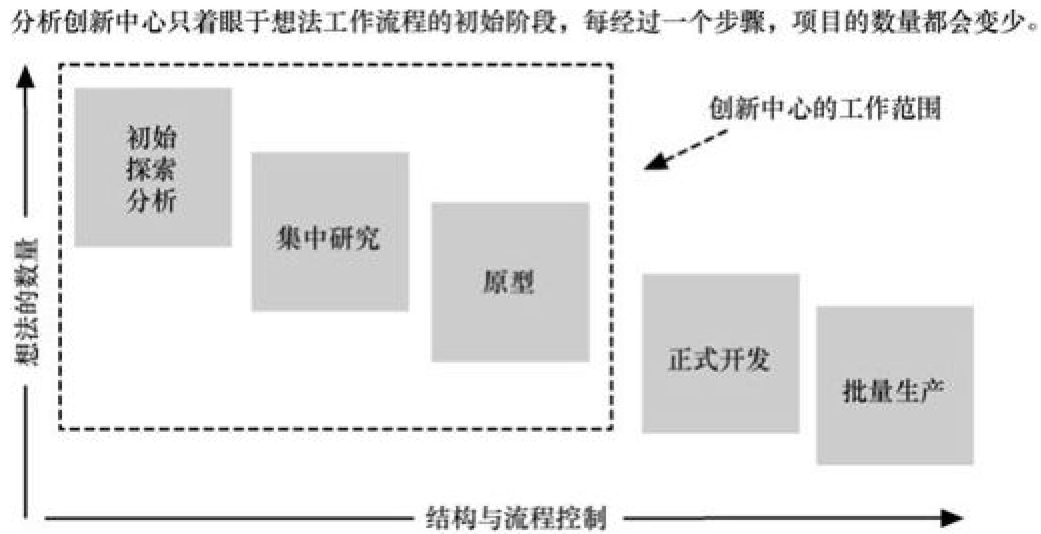
10.6.7 [illegible]

[illegible]

[illegible]

[illegible]

[illegible] 10-1 [illegible]



分析创新中心只着眼于想法工作流程的初始阶段，每经过一个步骤，项目的数量都会变少。
创新中心的工作范围
初始
探索
分析
集中研究
原型
正式开发
批量生产
想法的数量
结构与流程控制

[illegible]10-1 [illegible]

[illegible]

10.6.8

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] 4 [illegible] A [illegible] 4 [illegible] A [illegible] 4 [illegible] A [illegible]

[illegible]

10.7

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

[illegible]11[illegible]
[illegible]

[illegible]
iPhone[illegible]Wii[illegible]

[illegible]

[illegible]advanced analytics[illegible]
ThinkGeek.com[2][illegible]USB[illegible]

11.1 [illegible]

Silly Bandz[illegible]Jibbitz[illegible]Silly Bandz[illegible]Silly Bandz[illegible]Silly Bandz[illegible]Silly Bandz[illegible]Jibbitz[illegible]Crocs[illegible]Crocs[illegible]Jibbitz[illegible]Crocs[illegible]
Jibbitz[illegible]

[illegible]

[illegible]Crocs[illegible]
Jibbitz[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

11.1.1 Crocs[illegible]Jibbitz[illegible]

[illegible]Crocs[illegible]“[illegible]”[illegible]Crocs[illegible]

[illegible]Jibbitz[illegible]Jibbitz[illegible]Crocs[illegible]“[illegible]Crocs[illegible]Crocs[illegible]3[illegible]4[illegible]5[illegible]”

[illegible]Jibbitz[illegible]2005[illegible]Rich Schmelzer[illegible]Sheri Schmelzer[illegible]Jibbitz[illegible]2000[illegible]Crocs[illegible]

40[illegible]1976[illegible]SAS[illegible]SAS[illegible]“[illegible]SAS”[illegible][3][illegible]

11.1.2 [illegible]

[illegible]Jibbitz[illegible]Schmelzer[illegible]Crocs[illegible]Jibbitz[illegible]2000[illegible]Schmelzer[illegible]

[illegible]Jibbitz[illegible]Jibbitz[illegible]Schmelzer[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

11.2 [illegible]

[illegible]

■ [illegible]**1**[illegible]

■ [illegible]**2**[illegible]

■ [illegible]**3**[illegible]

[illegible]

11.2.1 [illegible]1[illegible]

[illegible]

[illegible]

[illegible]7[illegible]5[illegible]6[illegible]4[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1[illegible]

[illegible]15[illegible]16[illegible][4][illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]8[illegible]9[illegible]

2[illegible]

[illegible]4[illegible]

[illegible]20[illegible]90[illegible]

[illegible]

[illegible]40[illegible]50[illegible]100[illegible]MapReduce[illegible]4[illegible]MapReduce[illegible]

[illegible]MapReduce[illegible]

11.2.2 [illegible]2[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]20[illegible]90[illegible]Amazon[illegible]eBay[illegible]craigslist[illegible]LinkedIn[illegible]Facebook[illegible]

[illegible]GPS[illegible]1[illegible]

[illegible]2[illegible]100[illegible]

2[illegible]

[illegible]3[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

3[illegible]

[illegible]MapReduce[illegible]

[illegible]

[illegible]

[illegible]MapReduce[illegible]

[illegible]

[illegible]

[illegible]

11.2.3 [illegible]3[illegible]

[illegible]

[illegible]1[illegible]

[illegible]2[illegible]

[illegible]3[illegible]

[illegible]

1[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”

[illegible]“[illegible]”[illegible]

[illegible]

Stephen Covey“”[5]

3

2

[illegible]

[illegible]

[illegible]

[illegible]10[illegible]Facebook[illegible]

[illegible]

[illegible]Facebook[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

3[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

4[illegible]

[illegible]

[illegible]MapReduce[illegible]MapReduce[illegible]

[illegible]

11.3 [illegible]

[illegible]

■ [illegible]

■ [illegible]1[illegible]2[illegible]3[illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

■ [illegible]

[illegible]

[1] Merriam-Webster, "Innovation," www.merriam-webster.com/dictionary/innovation.

[2] Think Geek, www.thinkgeek.com/geektoys/cubegoodies/c208/.

[3] SAS, www.sas.com/company/about/history.html.

[4] [illegible] "Copernicus, Nicolaus," www.newworldencyclopedia.org/entry/Nicolaus_Copernicus[illegible]

[5] Stephen R. Covey, www.stephencovey.com/7habits/7habits-habit2.php.

MapReduce

[illegible]

[illegible]

TERADATA.

“[illegible]”

[illegible]

“[illegible]”

[illegible]

“[illegible]”

[illegible]

“[illegible]”

[illegible]